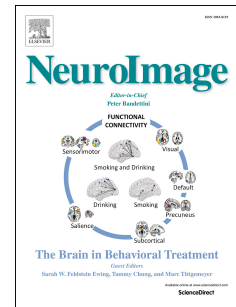


Accepted Manuscript

Evaluation of non-negative matrix factorization of grey matter in age prediction

Deepthi P. Varikuti, Sarah Genon, Aristeidis Sotiras, Holger Schwender, Felix Hoffstaedter, Kaustubh R. Patil, Christiane Jockwitz, Svenja Caspers, Susanne Moebus, Katrin Amunts, Christos Davatzikos, Simon B. Eickhoff



PII: S1053-8119(18)30192-7

DOI: [10.1016/j.neuroimage.2018.03.007](https://doi.org/10.1016/j.neuroimage.2018.03.007)

Reference: YNIMG 14776

To appear in: *NeuroImage*

Received Date: 3 January 2018

Revised Date: 28 February 2018

Accepted Date: 3 March 2018

Please cite this article as: Varikuti, D.P., Genon, S., Sotiras, A., Schwender, H., Hoffstaedter, F., Patil, K.R., Jockwitz, C., Caspers, S., Moebus, S., Amunts, K., Davatzikos, C., Eickhoff, S.B., Evaluation of non-negative matrix factorization of grey matter in age prediction, *NeuroImage* (2018), doi: 10.1016/j.neuroimage.2018.03.007.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Evaluation of non-negative matrix factorization of grey matter in age prediction

Deepthi P. Varikuti^{1,2,3}, Sarah Genon^{1,2,3*}, Aristeidis Sotiras⁴, Holger Schwender⁵,
 Felix Hoffstaedter^{1,2,3}, Kaustubh R. Patil^{1,2}, Christiane Jockwitz^{1,6,7}, Svenja Caspers^{1,6,7}, Susanne Moebus⁸,
 Katrin Amunts^{1,6}, Christos Davatzikos⁴, Simon B. Eickhoff^{1,2,3*}

¹Institute of Neuroscience and Medicine (INM-1, INM-7); Research Centre Juelich, Germany

²Institute of Systems Neuroscience, Medical Faculty; Heinrich-Heine University Düsseldorf, Germany

³Institute of Clinical Neuroscience and Medical Psychology; Heinrich-Heine University Düsseldorf, Düsseldorf, Germany

⁴Section for Biomedical Image Analysis, Center for Biomedical Image Computing and Analytics, University of Pennsylvania, Philadelphia, USA

⁵Mathematical Institute, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

⁶C. & O. Vogt Institute for Brain Research, Medical Faculty, Heinrich Heine University, Düsseldorf, Germany

⁷JARA-BRAIN, Juelich-Aachen Research Alliance, Juelich, Germany

⁸Institute of Medical Informatics, Biometry and Epidemiology, University of Duisburg-Essen, Essen, Germany.

*Both authors contributed equally to the work

Corresponding author:

Deepthi P. Varikuti

Institute of Neuroscience and Medicine (INM-7),

Research Center Juelich, D-52425 Juelich, Germany.

Email: deepthi.varikuti@gmail.com; S.Eickhoff@fz-juelich.de

Telephone +49-2461-61-96411; Fax: +49-2461-61-3483

SHORT TITLE: Age prediction using Non-negative matrix factorization

KEY WORDS: Non-negative matrix factorization, Structural MRI, Voxel-based morphometry, Age prediction, Dimensionality reduction, LASSO regression.

Abstract

The relationship between grey matter volume (GMV) patterns and age can be captured by multivariate pattern analysis, allowing prediction of individuals' age based on structural imaging. Raw data, voxel-wise GMV and non-sparse factorization (with Principal Component Analysis, PCA) show good performance but do not promote relatively localized brain components for post-hoc examinations. Here we evaluated a non-negative matrix factorization (NNMF) approach to provide a reduced, but also interpretable representation of GMV data in age prediction frameworks in healthy and clinical populations.

This examination was performed using three datasets: a multi-site cohort of life-span healthy adults, a single site cohort of older adults and clinical samples from the ADNI dataset with healthy subjects, participants with Mild Cognitive Impairment and patients with Alzheimer's disease (AD) subsamples. T1-weighted images were preprocessed with VBM8 standard settings to compute GMV values after normalization, segmentation and modulation for non-linear transformations only. Non-negative matrix factorization was computed on the GM voxel-wise values for a range of granularities (50 to 690 components) and LASSO (Least Absolute Shrinkage and Selection Operator) regression were used for age prediction. First, we compared the performance of our data compression procedure (i.e., NNMF) to various other approaches (i.e., uncompressed VBM data, PCA-based factorization and parcellation-based compression). We then investigated the impact of the granularity on the accuracy of age prediction, as well as the transferability of the factorization and model generalization across datasets. We finally validated our framework by examining age prediction in ADNI samples.

Our results showed that our framework favorably compares with other approaches. They also demonstrated that the NNMF based factorization derived from one dataset could be efficiently applied to compress VBM data of another dataset and that granularities between 300 and 500 components give an optimal representation for age prediction. In addition to the good performance in healthy subjects our framework provided relatively localized brain regions as the features contributing to the prediction, thereby offering further insights into structural changes due to brain aging. Finally, our validation in clinical populations showed that our framework is sensitive to deviance from normal structural variations in pathological aging.

1 Introduction

The structural dynamics of the human brain during adulthood is a highly complex process. Machine-learning algorithms have been used to capture the multivariate pattern of structural brain changes (Franke et al. 2010) that relate to age with a brain-based age prediction framework. By suggesting an age for any individual based on his/her brain's structural scan, such approaches can provide new insights into brain plasticity, into accelerating cerebral aging, as well as into the influence of several variables such as genes, pharmacological intervention and cognitive training in both healthy and clinical populations. Voxel based morphometry (VBM) is one of the most commonly used methods to measure grey matter volume (Good et al. 2001). It provides non-negative measures, which convey biologically meaningful information and capture brain changes related to age and pathology, as well as brain plasticity related to training (Good et al. 2001; Tisserand et al. 2002; May 2011). Previous studies have shown that machine-learning methods applied to VBM data allow prediction of brain age (Franke et al. 2010). In these studies, brain age was estimated by applying a support vector machine approach on the high dimensional voxels' data (Erus et al. 2015). However, in voxel-wise representation of structural data, features may convey redundant information and/or noise and may promote overfitting due to a higher number of features relative to the number of subjects (Guyon and Elisseeff 2003; Hua et al. 2009; Mwangi et al. 2014). To address this issue, Franke et al. (2010) examined brain age prediction based on the simple and fast application of the principle component analysis (PCA) to the data and subsequent brain age prediction with a relevance vector machine approach. This combination allowed them to predict the brain age with an absolute error of 5 years. Ever since, Franke et al. 2010's framework has been employed to investigate other concepts in relation to healthy aging (such as different age groups i.e., children and adolescents (Franke et al. 2012), gender differences (Franke et al. 2014)), as well as differences between healthy aging and various neurocognitive deviancies (such as cognitive impairments (Gaser et al. 2013) and psychiatric disorders (Koutsouleris et al. 2014)).

Most of the above-mentioned studies have implemented principle component analysis (PCA) to counter the curse of dimensionality associated with multivariate analysis of neuroimaging data (Franke et al. 2010; Franke et al. 2012; Franke et al. 2013; Liem et al. 2017). PCA decomposes the entire non-negative representation into a low rank approximation with a combination of positive and negative weights (Jolliffe 2002), which does not promote spatially localized components. Furthermore, the signed components within the PCA decomposition engage complex cancellations during the reconstruction of the original representation. Therefore, the use of

PCA-based dimensionality reduction on brain voxels hardly results in interpretable components, which can in turn prevent the interpretation of a predictive model based on PCA-derived components. Non-negative matrix factorization (NNMF) is an alternative decomposition method promoting relatively localized (spatial) representation that has gained more attention in the past years. NNMF can factorize a given dataset into low-ranking approximations capturing a parts-based representation (Lee and Seung 1999). The non-negativity constraint leads to only additive combinations of the components, which allows the factorization to reconstruct the original high dimensional data from the parts-based representation. As a result, NNMF provides a more interpretable factorization compared to standard decomposition approaches such as PCA and ICA (Independent Component Analysis) (Lee and Seung 1999; Sotiras et al. 2015). Recently, Sotiras et al. (2015) investigated the application of NNMF to neuroimaging data, by decomposing the structural MRI data with an extended version of NNMF, the orthonormal projective non-negative matrix factorization (OPNMF). This approach provided components that could be considered as a biologically more meaningful parts-based representation of the brain as compared to more standard approaches such as PCA and ICA. To note, OPNMF to some extent, generate bilaterally symmetric spatial features, despite being an unsupervised data-driven factorization approach (Sotiras et al. 2017). Accordingly, OPNMF promotes relatively localized (spatially) brain components for post-hoc examinations compared to the standard approaches. Hence, OPNMF could open new perspectives for dimensionality reduction of VBM data, in particular in a prediction framework. However, to the best of our knowledge, these perspectives have remained unexplored. Therefore, the current study aimed at examining the application of OPNMF to VBM data in a brain-age prediction framework. To note, we have used the term NNMF when denoting to the whole family of the technique, whereas, OPNMF when referring to the more the specific variant, which we have employed in this study.

We first compared the performance of OPNMF-based factorization to the performance of plain VBM data for age prediction. Then, in order to provide direct comparison with previous studies, we evaluated different strategies combining either PCA or OPNMF as a data compression approach with either LASSO or RVM as sparse regression models. In addition to the sparseness inducing methods described above, several parcellations of the human brain have been proposed in the last two decades (Eickhoff et al. 2017), which could potentially offer another efficient approach for data compression into relatively localized spatial units for age prediction. In particular, many whole-brain parcellations have been derived from voxels/vertex functional signal at rest (RS, e.g.: Bellec et al. 2010; Craddock et al. 2012; Gordon et al. 2016; Schaefer et al. 2017) and such RS-based

parcellation has been used for the compression of RS functional connectivity (RSFC) data in a brain age prediction framework (Liem et al. 2017). However, we assume that such a representation based on functional parcellation is, by nature, less optimal than a representation based on the structural properties of the voxels as used in the current VBM-based framework. To investigate this hypothesis, we compared the pattern of representation, as well as the prediction performance of our data reduction approach OPNMF capitalizing on structural covariance with an independent brain representation derived from resting-state functional data in healthy adults.

OPNMF is computationally more expensive than popular decomposition methods such as PCA (see methods). Nevertheless, transferring the factorization derived from one dataset onto another dataset could save this computational cost. Furthermore, using factorization from an independent dataset for training or testing a prediction model can assess the robustness of the model. We, therefore, evaluated the transferability of the OPNMF onto an unseen dataset, that is, we examined the transferability of the components derived from one dataset onto an independent (new) dataset, hence avoiding the time-consuming step of factorization in the new dataset. Importantly, transferring the already computed components onto a new dataset is particularly useful in clinical and research practices, as the datasets often come from different sites and scanners and may have different demographic characteristics. Recently, Liem et al. (2017) suggested that combining datasets from different protocols could reduce the bias of the predictive model towards the characteristics of a single protocol. Therefore, the effect of data acquisition and demographic heterogeneity on the transferability of the components is an important aspect to evaluate in the perspective of application of our framework in future studies. Here, we examined a dataset from a uniform protocol constituting older subjects (age range 55-76) vs. a heterogeneous multi-site dataset whose age range covers the adult life span (19-81, Fig 1A)). Thus, we assessed the performance of the prediction model trained on a dataset compressed using its own factorization, as well as, when this dataset was compressed based on an independent factorization (that is, when the dataset has been projected onto a factorization derived from a different dataset).

In addition, the difference in the sample characteristics of the two cohorts further offer the opportunity to investigate the extrapolation of the prediction model trained on one dataset onto an independent dataset. That is, in the present study, we investigated both, the transferability of the components among datasets and the generalization of the prediction among datasets, on the age prediction performance. Relatedly, one crucial

objective in age prediction is the identification of aging trajectories deviating from normal range, i.e., pathological aging. Previous studies have shown dramatic brain structural alterations in participants with deviations from healthy aging such as Mild Cognitive Impairment (MCI) and patients with pathological aging such as Alzheimer's Disease (AD) resulting in systematic overestimation of their age by an algorithm trained on healthy populations (Davatzikos et al. 2009; Gaser et al. 2013; Moradi et al. 2015). Therefore, as a validation of our framework for clinical research, we further evaluated its performance in age prediction of healthy and clinical samples from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database.

To sum up, in this study, we aimed to evaluate a new framework for brain-age prediction, which used dimensionality reduction of VBM data using OPNMF followed by a sparse regression model. In order to evaluate the advantages and limitations of this framework over the other approaches proposed in the previous studies, we compared the performance of our model with 1) model based on voxel-wise VBM data (uncompressed VBM data), 2) model based on PCA data reduction and 3) model based on data reduction based on RS-based parcellation. In the sake of reducing computational cost in future studies, we examined the transferability of the OPNMF between two independent datasets differing in demographic characteristic and acquisition protocols. Importantly, the localized properties of the components in our framework allowed us to explore brain regions contributing to the predictiveness in the healthy samples. Finally, we tested the performance of our prediction model on a clinical sample, in order to validate the predictive utility of our framework in clinical research.

2 Material and Methods

2.1 Sample characteristics and preprocessing

We used structural MRI data from two large, independent datasets. The first was obtained from the population-based 1000BRAINS study (Caspers et al. 2014) and represents a single-site assessment of 693 older adults (age: 55-75 years; 53% males) using the same imaging protocol for all subjects. The other “MIXED” dataset consists of 1,084 healthy adults (age: 18-81 years; 51% males) that were derived by pooling data from many different individual studies at various sites (Fig 1A; for further details see Supplementary methods). Furthermore, in order to validate our framework of age prediction on clinical data, we included a dataset from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). This ADNI dataset sample included 244 cognitively normal elderly subjects (HC, age: 55-90; 48% males), 64 mild cognitively impaired (MCI) subjects (age: 55-87; 60% males), and 163 Alzheimer's disease (AD) subjects (age: 56-91; 56% males), for further details see Supplementary methods.

Structural MRI data was preprocessed with the VBM8 toolbox (<http://www.neuro.uni-jena.de/vbm8>) to derive voxel-wise grey matter volumes for each subject of the two datasets using standard settings. T1-weighted structural brain images were normalized by the high-dimensional DARTEL normalization (Ashburner 2007) combined with tissue class segmentation and bias field correction. The normalized grey matter segments were modulated for non-linear transformations only and smoothed with an 8-mm FWHM Gaussian kernel. The local grey matter volumes (following adjustment of head size given that the affine part of the registration did not enter the modulation) were then extracted in a whole-brain grey matter mask (with a threshold of 0.2 to eliminate the voxels with partial volume effect (Ashburner et al. 1985)) and for each sample individually stored in a Number of subjects by Number of voxels matrix (with Number of voxels = 344,383). These matrices provided the input for the age-prediction model based on the full (uncompressed) VBM data as detailed in section 2.4 and the input to which matrix factorization (i.e., non-negative matrix factorization and principle component analysis) and resting-state (RS) based parcellation (see below) were applied.

2.2 Data reduction

2.2.1 Orthonormal projective Non-negative matrix factorization

We used the same orthonormal projective non-negative matrix factorization (OPNMF) approach (Yang et al.,

2007; Yang and Oja 2010) as described by Sotiras et al. 2015. OPNMF factorizes a data matrix 'X' into two non-negative sub matrices (W and H) representing the sparse components (the dictionary i.e., W of size, Number of voxels by Number of components) and the subject-specific loading coefficients (H of size, Number of components by Number of subjects) in the ensuing low-rank space, $\min_{W \geq 0, H \geq 0} \|X - WH\|_F^2$, which minimizes the squared Frobenius norm (i.e., reducing the reconstruction error), subject to the conditions $H = W^T X$ and $W^T W = I$ where, $\|\cdot\|_F^2$ referred to the squared Frobenius norm and I denotes the identity matrix.

To summarize the factorization process, W is first initialized through non-negative double singular value decomposition (NNSVD; cf. Boutsidis and Gallopoulos 2008)). Later, W is iteratively updated with the multiplicative update rule, until it converges to an optimum solution. The multiplicative update rule is modified as reported by Yang and Oja (2010), to satisfy the additional constraints of an orthonormal projection basis function, $W'_{ij} = W_{ij} \frac{(XX^T W)_{ij}}{(W W^T X X^T W)_{ij}}$, where, $i = 1 \dots \text{Number of voxels}$, $j = 1 \dots \text{Number of components}$. Finally, projecting X onto W to obtain a solution that minimizes the reconstruction error yields H. Following OPNMF, the VBM data are represented by two matrices, denoting the sparse components (W) and the corresponding subject-specific loading coefficients (H). The former (W) represent the latent structure in the data, the latter (H) represents the individual volumetric data in the low-rank space spanned by these components and provides the features for the age-prediction model. Of note, the highest possible OPNMF granularity is the lowest dimension of the input matrix (X) (which in our case is the number of the subjects in 1000BRAINS dataset being the smallest sample size). Accordingly, in this study, to explore the effects of different granularity, i.e., number of components, on prediction accuracy, we computed and evaluated compressions employing 50 to 690 components in steps of 25.

2.2.2 Principle component analysis (PCA)

PCA is one of the most commonly used dimensionality reduction techniques and, accordingly, has been used in previous studies examining age prediction based on structural MRI data. In order to provide a direct comparison of the OPNMF's performance with the previous investigations, we ran PCA on the voxel-wise VBM data by using the PCA function implemented in MATLAB 2014. This transformed the high-dimensional voxel wise data (i.e., X) into low-rank approximations using an orthogonal linear transformation. The resulting PCA based low-rank approximations represented the principle components of the data (computed by solving an eigenvalue

problem) arranged in descending order of the variance explained by each component (Jolliffe 2002). The subject-specific loading coefficients were obtained by projecting the high-dimensional voxel wise data onto the component space (eigenvectors, i.e., PCA based low-rank approximations) thereby providing the features for age-prediction model. Finally, we computed and evaluated the effect of PCA compressions on prediction accuracy, in the range of granularity aforementioned for OPNMF (i.e., 50 to 690 components in steps of 25.)

2.2.3 Resting-state (RS) based parcellation

Recently, Schaefer et al. 2017 reported a parcellation based on RS fMRI providing neurobiologically-valid brain parcels by capitalizing on a new hybrid approach integrating the local gradient approach for boundary-mapping with a global similarity approach. As this atlas does not cover subcortical and cerebellar structures, we added these from another widely used RS fMRI parcellation (BASC, Bellec et al. 2010). This resulted in a whole brain parcellation of 470 parcels that was used here as an alternative dimensionality reduction approach for VBM data. For each subject, an average grey matter volume within each parcel was computed and used as inputs for the age-prediction model.

2.3 Sparse regression model

We primarily used LASSO (Least Absolute Shrinkage and Selection Operator) for learning a (sparse) linear regression model predicting the subjects' age from their structural data as compressed in the loading coefficients (as implemented in the 'glmnet' package, <https://www.jstatsoft.org/article/view/v033i01> (Tibshirani 1996)). LASSO regulates the parameters (alpha and lambda) to optimize the sparsity and the complexity of the regression model to improve the performance (i.e., prediction accuracy) and interpretability of the model (Zou and Hastie 2005; Zhang and Huang 2008). An inner loop was incorporated to optimize the hyper-parameter (lambda). LASSO with alpha set to 0.99 and lambda that gives minimum mean cross validation error of the inner loop was employed for predicting the age in our study.

As an alternative approach to LASSO, Relevance Vector Machine (RVM; Tipping 2001) has been commonly implemented by the previous studies exploring prediction of age using structural MRI. Therefore, we in this study performed an additional comparison between LASSO and RVM regression models. For doing so, statistical learning of the sparse regression model employing RVM was implemented using the SparseBayes package (<http://www.miketipping.com/index.htm>). This approach uses a probabilistic Bayesian framework with

specific priors over the parameters, which favors sparse prediction model. The algorithm iteratively and automatically optimizes the parameters and hyper parameters, hence reducing prior control on the parameters. As kernel, we chose a multivariate zero-centered Gaussian with standard deviation estimated by the algorithm. This RVM implementation from the SparseBayes package has been shown to improve the initialization procedure, which maximizes the likelihood function and hence accelerates the procedure (Tipping and Faul 2003).

2.4 Prediction analyses

Previous studies of age prediction from MRI data in life span cohorts have used linear regression model (Franke et al. 2010; Franke et al. 2012; Gaser et al. 2013; Mwangi et al. 2013; Franke et al. 2014; Koutsouleris et al. 2014; Erus et al. 2015; Liem et al. 2017). For the sake of comparability, we likewise used a (sparse) linear regression model for predicting the subjects' age from their structural data as compressed in the loading coefficients. Furthermore, in the present study, combining sparse decomposition method with a sparse regression model came with the advantage of providing an anatomically well interpretable model for estimating age based on a limited number of spatially compact structural features.

Model generalization was evaluated by 10-fold cross-validation. That is, the dataset was randomly split in ten equal parts that each, in turn, served as the test set for the model fitted on the remaining 9/10th of the data. To reduce dependency on the cross-validation split, this procedure was replicated 100 times. Splitting the dataset into ten equal parts has been initiated every time within each repetition, which allowed us to train the model on different training samples in each repetition. Prediction accuracy was quantified by the mean absolute deviation (across subjects) between real age and predicted age (averaged across repetitions), and also, the correlation between the real age and that average of the predicted (across repetitions) in previously unseen subjects from their VBM data.

2.5 Assessed prediction approaches

We note that performing OPNMF only on the training dataset in each cross-validation fold would be computationally expensive and hence practically infeasible. But prediction performed on loading coefficients obtained from the OPNMF decompositions including the entire sample (including the 1/10th that is denoted the test-set in the respective fold) could bias the cross-validation towards overly optimistic performance estimates.

Given this consideration would also hold for all future uses of our approach, we were particularly interested in investigating whether components derived from one dataset were also good encoders (representative) for the structural features of another dataset.

Consequently, we performed cross-validation analyses using the loading coefficients derived from OPNMF of that particular dataset as (an overly optimistic) reference (Fig 2: A&B) but importantly focused on assessing the possibility to predict subjects' age after projecting the raw VBM data on the component space estimated for the other dataset (Fig 2 C&D). That is, we derived the OPNMF components of the 1000BRAINS dataset, and performed cross-validation within the MIXED dataset projected onto the components estimated from the 1000BRAINS. This approach has the advantage that the subjects in the test set were truly independent and have not been involved in any prior processing steps. In addition, we could investigate the effects of dataset (in-) homogeneity, as the 1000BRAINS data comes from a single-site study with uniform protocol, whereas the MIXED dataset was deliberately chosen to be very heterogeneous. We would thus expect that the components derived for the MIXED dataset show a better generalization than those from the 1000BRAINS dataset, i.e., projecting the 1000BRAINS data onto the components from the MIXED dataset will yield prediction models that are closer in performance to the (optimistically biased) analysis with the projection of the data components derived from the MIXED data itself.

Later, we also tested whether the actual models transfer between datasets by deriving the OPNMF components in one dataset (e.g., MIXED), fitting the sparse regression model in that same dataset (MIXED), projecting the other dataset (1000BRAINS) onto the factorization of the former (MIXED) and applying the prediction model trained on that (MIXED) data (Fig 2: E, F&G). In this context, we note that the 1000BRAINS dataset has a more restricted age-range (55-76) than the MIXED dataset (18-81). Therefore, we evaluated the model transfer between the portion of MIXED subjects corresponding to 1000BRAINS age range (55-75; i.e., OldMIXED), see Fig 2: E&F. The computational times for each prediction approach at different levels of granularity are reported in supplementary material (Table S3).

Finally, our age estimation framework was validated using the ADNI database. Here we compared the estimated BrainAGE between healthy controls (HC=, subjects with mild cognitive impairment (MCI) and Alzheimer's disease (AD), given that apparent older brains have been previously demonstrated in the latter two groups

(Franke et al. 2010; Franke and Gaser 2012). As AD patients sample mainly consisted of older subjects, the prediction model was trained on the aforementioned samples of elderly subjects (i.e., 1000BRAINS or MIXED_55-90). In detail, all data (training sample and ADNI) were projected onto the factorization derived from the respective training sample (either the 1000BRAINS or MIXED). The model was trained on the each of the training sample (1000BRAINS or MIXED_55-90 (i.e., Subjects above 55 years from MIXED dataset)) and then evaluated it in the ADNI data. In line with previous studies (Davatzikos et al. 2008; Franke et al. 2010; Franke and Gaser 2012; Moradi et al. 2015), we hypothesized, that for the ADNI controls, the brain age gap estimation (BrainAGE), i.e., the difference between the predicted age and the chronological age, should be centered around zero. In turn, MCI subjects and in particular AD patients were expected to show an increased BrainAGE score.

2.6 Identification of the regions influencing the prediction

As noted above, combining a sparse decomposition yielded compact regional modes with a sparse regression model (LASSO) has the advantage of providing regionally specific relevant features. As a final step allowing the neurobiological interpretation of our age-prediction model, we identified those parts of the brain that underpinned the reported predictions. As we performed 100 replications of a 10-fold cross-validation, in total 1000 models were computed per granularity (number of components). We then quantified the contribution of each component by the fraction of these 1000 models in which the coefficients assigned by the predictive model for the respective component was non-zero. The components that contributed in at least 95% of all models were identified as the components that were robustly part of the predictive model (Fig 3 (5)). Concretely, we first identified the components consistently contributing to the prediction as defined by non-zero beta value in 95% of the models. Second, the components were mapped to the brain space at each respective level of granularity. That is, we built a “contributor map” at each level of granularity, in which the voxel values represent their (binary) contribution (Fig 3 (5)). Combining those maps (by summing up the values) resulted in a contributor “summary” map in which a non-zero value represents a contribution in at least 95% of the 1000 prediction models) and higher value represent more overlap across different granularities (Fig 3 (6)). As we found that prediction performance stabilizes after around a granularity of 300 components (Fig 6), only the contributor maps of granularity > 300 components were merged into a summary map. Given that the relationship between the OPNMF components at high granularities could be hierarchically inconsistent, this approach yielded a

higher effective resolution of the relevant brain areas than the actual granularity of the factorizations itself, and hereby alleviated the reliance of the spatial inference on any particular set of components.

3 Results

3.1 Brain age estimation using the uncompressed VBM data

Training LASSO models on the full, i.e., uncompressed voxel-wise VBM data allowed to predict the age of previously unseen subjects with relatively high accuracy. For the 1000BRAINS data, the mean absolute error (MAE) between real and predicted age of the test set was 3.4 years. While for the MIXED dataset, the MAE was 4.9 years. While these numbers compare favorably with previous reports, Fig 4 illustrated the critical drawback of using sparse regression models on voxel-wise data. That is, isolated voxels scattered across the brain were selected as relevant features by the prediction model. In addition to being computationally prohibitive, the ensuing models are basically uninterpretable as the predictions were driven by individual voxels (Fig 3).

3.2 Compression of Brain age estimation using different compression methods and sparse regression models

Fig 5 illustrated the performance of each of the four combinations of approaches. Across different cross-validation approaches (Fig 5A), OPNMF either slightly outperformed or remained analogous to PCA, especially at higher level of granularity. In particular, when the factorization has been transferred from one dataset to another dataset (Fig 5A: plots on the right compared with plots on the left), OPNMF reported more accurate age prediction with stable performance across different levels of granularity compared to PCA. Thus, we could infer from our results that OPNMF derived from one dataset could provide a better representation of the structural data of an independent dataset than PCA. With respect to the sparse regression approach, LASSO and RVM resulted in comparable cross-validation accuracies, but LASSO was shown to yield superior performance when predicting age across samples (Fig 5B), irrespective of the employed factorization. Additionally, the application of LASSO together with OPNMF performed better than the other combinations in most of the scenarios, supporting the combination of LASSO with OPNMF for age prediction analyses. Accordingly, we focused on investigating the brain age prediction using LASSO sparse regression model, in the subsequent analyses (such as, comparison of OPNMF with a previous RS-parcellation, examination of the OPNMF transferability, identification of the relatively localized features contributing to the prediction analyses in healthy datasets and

finally validation of our framework (i.e., combination of OPNMF with sparse regression model) in a clinical dataset).

As suggested by our comparative analyses, based on the prediction accuracies reported by the LASSO (sparse) regression model, we compared the performance of the OPNMF factorization with a RS-parcellation of the brain (Bellec et al. 2010; Schaefer et al. 2017; cf Section 2.2.3). As illustrated in Fig 6, at comparative levels of granularity (i.e., 475 OPNMF factors vs. 470 brain parcels), the age prediction model tended to be more accurate when the data have been compressed with OPNMF than when the data have been compressed based on an independent representation derived from RS fMRI signal. Nevertheless, it has to be noted that the latter also compressed data into localized brain parcels, which by itself showed surprisingly good performance, suggesting that different spatial representations into local components can be efficient (see discussion). In the scope of the current study, altogether, our preliminary comparative analyses supported the use of OPNMF for data compression in an age prediction framework.

3.3 Influence of different datasets on the OPNMF & age prediction

As previously reported, OPNMF provided sparse and spatially compact components, which essentially reflect local structural covariance (Fig 1B). While not the primary focus of this work, we noted that labeling each grey-matter voxel by the most strongly reflected component, provided a map of the human brain that in many aspects seemed to resemble those derived from other modalities. Across the whole range of granularity, although there seems to be a decent agreement, these maps were slightly different between both investigated datasets (1000BRAINS and MIXED) as reflected in the adjusted rand index (aRI, Fig 1C). This latter quantifies the similarity between the clusters (Hubert and Arabie 1985; Santos and Embrechts 2009)) between the respective parcellation and can range between +1 and -1, with 1 reflecting perfect spatial correspondence, 0 indicating spatial agreement with certain probability, and smaller than 0 representing disagreement which is worse than contingency (Hubert and Arabie 1985)). However and more importantly, both factorization (1000BRAINS and MIXED) at the similar level of granularity (i.e., 475 granules) showed good convergence with the RS-parcellation (470 parcels) with, respectively, $aRI = 0.28$ and $aRI = 0.29$ (Fig 5).

3.4 Brain age estimation using the OPNMF-compressed VBM data

Considering the models based on the loading coefficients for components derived from the (full) data of the same sample (rendering data compression not independent from the latter cross-validation), several important

observations can be made. First, while very low-rank approximations only yielded poor prediction accuracy, the mean absolute error (MAE) quickly declines with increasing granularity, i.e., higher number of components. Once the number of components passes approximately 300 – 400 (Fig 6), however, prediction accuracy seems to remain stable or at best improve asymptotically.

3.4.1 Model validation within the same dataset using a cross-validation approach

In details, these unbiased models yielded an overall MAE of 3.6 years (males: 3.7 and females: 3.6) and an overall correlation of 0.65 (male: 0.62 and female: 0.61) between real and predicted age in the 1000BRAINS data (using components derived from the MIXED dataset; Table S1 & Table S2). For the MIXED data, we found an overall MAE of 6.1 years and a correlation of 0.88 (MAE of 6 and $r = 0.88$ in the males and MAE of 6.3 and $r = 0.86$ in the females) when using components derived from the 1000BRAINS dataset (see Table S1 & Table S2 for detailed results; also see the supplementary material for the discussion on association between the two measures (i.e., MAE and correlation) of the performance (Fig S6)). Further examining the prediction performance across the different scanning sites forming the MIXED dataset (16 different sites) revealed that in most of them (14 sites) the MAE varied between 5 to 7 years and the MAE from the two other sites were 4 years and 9 years (of note, the scanning protocols used in these latter remained analogous to the 14 other sites, i.e., we did not note any specific technical factor accounting for the differences in prediction accuracies). Overall, these results showed the stability of our prediction framework across genders and scanning sites.

3.4.2 Model validation with prediction in independent datasets

Transfer of the whole pipeline (factorization and model training) was evaluated by predicting the age of the subjects in the respective other, independent sample (Fig 6B & Fig 7B). In our study, transferability of the prediction model was evaluated in two different aspects, extrapolation of the prediction model onto an independent dataset, which differs in subjects' demographic characteristics, such as age, and then onto an independent dataset, which differs in scanner protocols. In the context of dataset (in-) homogeneity from different age groups, models trained on broader age range of the heterogeneous dataset (MIXED) showed reduced precision of age prediction in an independent dataset (1000BRAINS) while the model trained on a restricted age range for this restricted heterogeneous dataset (OldMIXED) was more accurate in predicting the age of the latter independent dataset (1000BRAINS). In the context of dataset (in-) homogeneity from different protocols, our results surprisingly showed that models trained on single-site study (1000BRAINS) also

performed efficiently, when predicting the age of highly heterogeneous dataset (OldMIXED). In contrast, models trained on the 1000BRAINS data consisting exclusively of older subjects showed a very poor performance when trying to predict age of the younger subjects in the MIXED sample (Fig S2). While the model correctly predicted the young subjects to be younger than the young examples in the training set, it was grossly inaccurate in predicting how much younger they actually were. Put pointedly, having no information about how a 20-year old brain looks like, a model trained on subjects aged between 55 and 76 can only derive that the subject in question should be younger than the youngest it has seen in the training data (Fig S2). Thus, testing for generalization of the model to an independent dataset showed good prediction accuracy for the subjects within the training sample's age range (Fig 6B & Fig 7B), but also indicated that the prediction model cannot extrapolate to the subjects whose age is (far) beyond the training samples age distribution (Fig S2).

As ultimately one main application of our framework will be research in clinical populations, we also tested our framework in the ADNI dataset. Here, the mean BrainAGE scores (reflecting, for each subject, the discrepancy between brain-based estimated age and chronological age) was zero in the healthy control group (for models trained on either the 1000BRAINS or the MIXED_55-90 datasets). In contrast, BrainAGE scores were 6.2 years (for models trained on 1000BRAINS) and 5.4 years (for model trained on MIXED_55-90) in the MCI group, indicating that these subjects' brains looked about 5-6 years older. Finally, the BrainAGE scores reached 8.5 years (for models trained on 1000BRAINS) and 10.7 years (for models trained on MIXED_55-90) in the group of patients diagnosed with Alzheimer's disease. These results, illustrated in Fig 8 demonstrated that our framework can accurately capture the range of normal structural variation relating to age in healthy subjects and building on this normal range, captures dramatic deviance in both patients with Mild Cognitive Impairment and patients with Alzheimer's disease.

3.5 Identification of the regions influencing the prediction

The framework we examined in this paper, i.e., applying a sparse regression model onto the sparse decompositions, should yield rather confined and hence neurobiologically interpretable maps of brain regions contributing to the age prediction. In more detail, as previously noted, the OPNMF components themselves were circumscribed rather than representing a mixture of voxel-wise positive and negative weights as would be the case for PCA (cf. Sotiras et al., (2015)). LASSO then selected a small number of these spatially confined components for the actual prediction. This allowed us to identify, which brain regions consistently contributed

to the age estimation. As demonstrated in Fig S7, many of the regions that were selected by the models are also the regions that show the highest correlation with age. This suggests that, to some extent, the observed contributors are consistent with age-related GMV changes.

As illustrated in Fig 9, more brain regions were engaged in estimating age in the MIXED as compared to the 1000BRAINS dataset, which could be expected given the much broader age distribution. More specifically, the regions contributing to the prediction model in the 1000BRAINS cohort (older subjects) included regions around the central sulcus, the inferior temporal cortex, the occipital and posterior temporal cortices and area 44. Regions contributing to the predictions in this older adult cohort also included bilateral midline areas such as, the superior medial frontal gyrus, the medial fronto-orbital regions, the anterior and middle cingulate cortices and the retrosplenial cortex. Furthermore, the pattern of regions weighting in the prediction model in this cohort further included bilateral subcortical regions such as the thalamus, the basal ganglia and the posterior hippocampus, as well as the bilateral cerebellum. On the lateral surface, the pattern included regions in prefrontal regions (frontal areas anterior to the precentral gyrus), orbitofrontal regions and temporal poles. In contrast, the pattern of regions for age prediction in the MIXED dataset (heterogeneous dataset covering the whole life span) was less spatially specific, covering most of the brain lateral surface bilaterally (including for example the whole bilateral middle and superior frontal gyri, as well as the bilateral posterior superior and inferior parietal cortices), almost the entire medial structures, and, the bilateral anterior hippocampus and amygdala. In other words, the prediction models of age in this heterogeneous dataset built on most of the brain regions.

3.6 Supplementary Analysis

Spatial smoothing on the VBM data promotes homogeneity of the data by attenuating small differences between individuals. In turn, age prediction may rely on those subtle effects. Thus, we also evaluated, whether the subjects' age could be predicted based on the unsmoothed VBM. As could be expected from the aggregation of individual voxels into components, refraining from smoothing prior to projection resulted in highly similar results as shown above for the smoothed data (see Supplement, Table S4 & Fig S1).

4 Discussions

In this study, we showed that non-negative sparse coding through the combination of data compression using OPNMF with LASSO regression could predict age of previously unseen subjects in an unbiased manner from structural neuroimaging data. Several key observations emerged from this work. i) The precision of age prediction compares well to that based on uncompressed, i.e., voxel-wise VBM data and to that based on non-sparse factorization (PCA). ii) Even though the components estimated for the two datasets differed from each other, the (unbiased) prediction accuracy after projection onto the respective other set of components is only slightly worse than the (biased) accuracy obtained when performing factorization of the entire dataset that was later used for cross-validation. iii) OPNMF-based brain partitions show some convergence with an independent parcellation based on resting-state (RS) fMRI, but the former gave slightly better prediction performance iv) Finally, in contrast to approaches used in previous age prediction studies, combination of data compression using OPNMF with sparse (LASSO) regression yields a superior interpretability of the weight maps allowing interpretations about the mechanisms underlying the prediction.

4.1 Prediction from uncompressed VBM data

Our results showed that age prediction of unseen subjects using the full (uncompressed) VBM data reported only slightly better prediction accuracies than one based on the (OPNMF) compressed (Table S1). This comparable level of performance for compressed and uncompressed data has also been observed in previous brain age studies employing PCA compression (Franke et al. 2010; Liem et al. 2017). However, in the current study, predicting age using sparse regularization (LASSO) prediction model on uncompressed VBM data is highly inefficient in terms of memory usage, especially for large datasets (for example, MIXED dataset with 1,084 subjects). In particular, the memory load of this high-dimensional approach (>700 subjects x 344383 voxels) only allowed a 3-fold cross-validation on a high-performance server. While high dimensional voxel wise data could also lead to overfitting of prediction model, due to the larger number of features than subjects (i.e., several models potentially could fit the same data), comprehensively investigating this issue was not possible in the present study due to the computational limitations. Beside this still open issue, the recent availability of MRI data in very large sample sizes, i.e., for thousands of subjects (e.g., (Miller et al. 2016)) and the growing interest for the prediction of phenotype or behavioral measures from MRI data, dramatically underpin the need of dimensionality reduction preserving prediction accuracy (Davatzikos 2016).

The key limitation of voxel-wise analysis, however, is the poor interpretability of the relevant features. As shown in Fig 4, the sparse regression model on the voxel-wise data in our study highlighted isolated voxels scattered over the brain as relevant features for predicting subjects' age. Nevertheless, the individual anatomical correspondence of a particular voxel chosen by the prediction model, can be variable across subjects (Davatzikos 2004). In addition, LASSO regression is known to perform reliable feature selection, providing that the features have followed "irrepresentable condition" (Zhao and Yu 2006). That is, features should be independent of each other in order to obtain reliable outcomes. Therefore, when LASSO is applied to voxel-wise VBM data, the isolated voxels from the highly correlated voxel-wise VBM data contributing to the prediction cannot really be interpreted. In other words, voxel-wise sparse regression models pose a decoding problem (Kampa et al. 2014). Thus, the poor interpretability of prediction models based on raw VBM data (Lakkaraju et al. 2016), in addition to their computational costs, advocate for data compression, ideally with a factorization approach that offers interpretability of the representations such as the current implementation of OPNMF on VBM data for prediction of brain age.

4.2 Compression of Brain age estimation using different compression methods with various sparse regression models

When comparing the performance of OPNMF with PCA, particularly at higher level of granularity, our results demonstrated that OPNMF either slightly outperformed or remained analogous to PCA. Any data reduction procedure aims to address the curse of dimensionality without any loss of information. In this context, both PCA and OPNMF provide low rank approximations representing the most influential structure within the original data, however, each decomposition method captures different aspects of the similar information (PCA captures the components with the most variance explained across the dataset, while OPNMF captures the spatially more localized components that consistently co-vary across the dataset), leading in the present study to comparable performance of both approaches in age prediction. Importantly, our results also further showed that OPNMF provided more stable performance at high granularities (> 200), when compression is transferred across datasets (Fig 5A: cf. right vs. left plots). This finding confirms previous hypotheses that the 'projectivity' of OPNMF supports the efficient transferability of the factorization onto a new unseen dataset (Yuan et al. 2007). Therefore, we would argue that OPNMF, compared to PCA, enhances the generalizability of the low rank approximations

onto an independent dataset. Thus, OPNMF not only promotes relatively localized brain representation, but also yields more generalizable low-rank approximation than PCA.

Our evaluation further revealed that LASSO regularization performed either similarly or slightly better than RVM. Both LASSO and RVM yield sparse regression models with the advantage of performing feature selection by capitalizing only on the features that improve the prediction accuracy and allow comparable accuracies. However, an additional argument for the use of LASSO, this model allows the selection of the regularization parameter. Hence, LASSO optimizes the trade-off between stability and interpretability of the prediction model (i.e., optimizing the sparsity) by tuning the regularization parameter (i.e., α), which linearly combines the L1 and L2 penalties (cf. Zou and Hastie 2005 for more technical details). Therefore, the LASSO regression model can convert the sparse regression model into a purely non-sparse model (using the elastic net regularization model) and can therefore be considered as a relatively more flexible regression model than RVM. Furthermore, Bunea et al. 2011 demonstrated that LASSO could be implemented in many conditions including when the feature size is exceeding the sample size for prediction analyses. Indirect support for this property of LASSO can be seen in top left plot of Fig 5B, in which LASSO works particularly better than RVM after crossing the granularity level of 250 (with the training sample size approximating 230 subjects in this specific case). As reduction techniques have shown best prediction accuracies at higher level of granularity and given previous considerations, we focused on LASSO for subsequent prediction analyses.

4.3 Compression of OPNMF with resting-state based brain parcellation (RS-parcellation)

The well above chance level ($\approx .30$) adjusted rand index between the RS-parcellation and the OPNMF indicates that the spatial representations derived from OPNMF based on structural covariance converge well with the spatial representation derived from resting-state functional signal in healthy adults. Of note, the used RS-parcellations have both been extensively evaluated in their respective studies, namely with regards to stability and convergence with histological mapping and alternative parcellations (Bellec et al. 2010; Schaefer et al. 2017). The RS-parcellations also define spatially homogenous regions, suggesting that the structural representations identified here capture segregated patterns of brain functional organization (Sporns 2013). Thus, the similarity between the brain partitions derived from OPNMF and the “optimized” RS-parcellation that we found in the current study allows us to assume that our OPNMF brain partitions have some biological validity. Similar observations have been reported recently by Sotiras et al. 2017 who showed that at low granularity (<

60), the components derived by OPNMF resembles previously evidenced functional brain networks. Together, these findings thus suggest that OPNMF of VBM data to some extent captures meaningful patterns of brain functional organization, both at the network and areal level.

While OPNMF-based factorization and RS-parcellation showing good convergence, they did not show a perfect agreement. This is in line with the few multi-modal mapping studies showing that brain maps from different features (such as structure and function) converge towards similar brain partition schemes, but also suggesting that each feature targeting a specific aspect of the brain tissue, each feature can capture a unique aspect of brain organization (Kelly et al. 2012; Genon et al. 2016; Glasser et al. 2016; Genon et al. 2017). In other words, different features (i.e. modalities) are to some extent sensitive to different aspects of brain organization (for a more detailed discussion see Eickhoff et al. 2017). From the perspective of data compression, the most efficient partitions should thus come from the same modality. And indeed, RS-parcellations provides more homogeneous parcels when assessing resting-state images than histologically defined brain regions (Craddock et al., 2012). This leaves the question, whether the amount of transferable information is still sufficient for a useful representation. Our results also provided evidence that this is the case by showing that a more accurate age prediction model is built from VBM data when this data is compressed directly as compared to representing it based on a functional parcellation of the brain (Fig 5) even if the latter yields very good accuracies. Overall, in the context of multivariate pattern analysis, we suggest that brain parcellation derived from one modality is transferable to another modality for data reduction even though it does not reach within-modality performance.

4.4 Influence of different datasets on the OPNMF

Despite the brain topographical pattern of the OPNMF components derived from the two different datasets show similar convergence with the independent RS-parcellation and general good agreement between them, they are not perfectly similar (Fig 1B & 1C; Fig S5). However, it has to be noted that the similarity between the factorizations derived from the two datasets has been measured at a level of granularity that does not favor reproducibility (even between datasets which are age, gender and site matched) according to previous work (Sotiras et al. 2015; Sotiras et al. 2017; Fig S5). As the granularity increases, the resolution of structural covariance increases resulting in a finer representation of covariance patterns, but that are, in turn, more influenced by covariance trends specific to the dataset used. Thus, we assumed that the difference in sample

characteristics between the two cohorts could explain the slight differences in the brain topographical pattern of the factorizations.

4.5 Impact of granularity on age prediction

Importantly, our study showed that despite the fact that reproducibility may decrease at high level of granularity (Sotiras et al. 2017), prediction performance did increase as granularity increases (as previously suggested by Sotiras et al., 2015). Our results demonstrated that when the number of components reaches approximately 300-400, prediction accuracy remains largely stable, particularly when the factorization is derived from the same dataset. However, when the factorization is derived from an independent dataset, a somewhat higher granularity (i.e., a few more than 400 components) might be required to reach stability. At a level of granularity around 300-400 components, OPNMF seemed to factorize the entire voxel-wise data into efficient subdivisions, which allowed the LASSO regression model to capture only the relevant features (i.e., ~116 features when predicting MIXED sample and ~52 features when predicting 1000BRAINS sample) and ignore the unnecessary/noisy features relatively better than at coarser granularity. Our finding converges with the study of Franke et al. (2010), in which the data compression was performed using PCA. These authors found that the lowest mean absolute error of the prediction analysis was reached at around 350 components. Of note, this level of granularity (or factorization) seems also convergent with the range of subdivisions of the brain that emerged as stable in functional MRI data, which lies between 200 and 500 parcels (Tucholka et al. 2008; Thirion et al. 2014; Gordon et al. 2016, Schaefer et al., in press). We could assume that a lower level of subdivision (i.e., < ~200 components) provides less homogeneous regions (i.e., regions mixing different functional and structural properties, cf. Eickhoff et al. 2017), while a higher level of subdivisions (>~ 500 components) might spatially narrow the components but without importantly improving the homogeneity within regions. Thus, the current study suggests that a factorization of VBM data into 300 to 500 components optimally organizes voxel-wise structural data into homogeneous brain regions for age prediction.

4.6 Model validation within the same dataset using a cross-validation approach

In the model validation within the same dataset, our study showed that the performance of the brain age prediction using the framework of non-negative sparse coding (i.e., non-negative matrix factorization with LASSO regression model) was similar to the prediction accuracy found in previous studies (Franke et al. 2010; Liem et al. 2017). It is important to note that predicting the age of subjects compressed using the components

derived from the same dataset violates the test set independence. Even though the subjects in the test dataset were separated from the training dataset at the prediction level, the used factorization reflects the best factorization of the entire dataset, including the test dataset (Yuan et al. 2007; Liu et al. 2010). In other words, the test dataset cannot be considered as strictly unseen because the test data has been “optimally spatially organized” with its own factorization scheme. Hence, performing the brain age estimations on the dataset compressed using the same dataset’s factorization facilitates optimistic predictions. Therefore, in this study we compared the performance of the proposed prediction framework with this later over-optimal approach.

Our results demonstrated that the LASSO regression, when applied on the dataset compressed with components derived from an independent dataset, estimated the brain age with a precision comparable to that achieved when compressing the dataset with its own factorization. This finding confirms the previous literature arguing that the ‘projectivity’ of OPNMF allows the efficient transferability of the factorization onto a new unseen dataset (Yuan et al. 2007). Our results showed that the differences between the factorizations derived from the two datasets (cf. section 4.3) did not influence the prediction of brain age when transferring the components onto the other dataset. Furthermore, our supplementary results (Fig S3) illustrated that the pattern of regions selected by the prediction approach remained similar when the factorization was derived from another dataset. That is, the prediction model recollected the same anatomical regions regardless of which factorization scheme was applied. Again, this pattern of findings converges with what has been previously observed in data reduction of fMRI data for subsequent functional connectivity analyses. Those parcellation studies have observed that at an optimal resolution, parcellation from one dataset can provide a relevant spatial representation of the functional signal in other datasets, despite the topographical pattern of the parcellation between the datasets being different (Bellec et al. 2010; Finn et al. 2015; Gordon et al. 2016). Similarly, OPNMF factorization based on a different dataset did not prevent an optimal compression of the data for age prediction or the selection of the relevant (anatomical) features. Thus, overall, our results demonstrated that despite the fact that factorization results from different datasets may comprise slightly different spatial components, any of the stable factorizations offers an efficient data compression for prediction analyses.

4.7 Model validation on an independent dataset

In the context of dataset (in-) homogeneity from different protocols, we observed that the prediction model extrapolated quite well to an independent dataset (Fig 6B & Fig 7B: Top left & bottom left plots). Firstly, model

trained on a highly heterogeneous dataset, better predicted the subjects' age in an independent dataset. Thus, our study supported Liem et al. 2017's recent suggestions that merging datasets from multiple protocols could avoid fitting the model to the characteristics of a particular scanner protocol. In other words, heterogeneous datasets allow the model to encounter a wider range of variations, helping it to disentangle non-relevant inter-individual variations from relevant variations for prediction. Surprisingly, model trained on single-site study also performed efficiently, when predicting the age of highly heterogeneous dataset (OldMIXED). To note, the single-site study consists of 693 subjects between 55 to 75 years (Fig 1A). Therefore, the model trained on this dataset encountered a wide range of variation at each age point. We suppose that this exposure to wide range of variation might have allowed overcoming the scanner effects with a robust regression model. Thus, we would recommend to train a given prediction model on a heterogeneous dataset (either with multi-sited examples or with multiple examples, ideally both) to ensure that true relevant variations are learned, which in turn may support good prediction performance. Importantly, the two cohorts also differed in their age distribution. Therefore, in addition to the generalizability over different protocols, these datasets also allowed us to evaluate the generalization of the prediction model over different age distributions (Fig 6B & Fig 7B: Top row plots). Not unexpectedly, models trained on restricted age range of the heterogeneous dataset (OldMIXED) provided better age prediction for test sample coming from an independent dataset within the age range of the training sample (1000BRAINS) when compared to the model trained on broader age range (MIXED). Again to be expected, models trained on narrow age range single-site study (1000BRAINS) failed to predict the age of subjects (MIXED) that were out of the training sample's age range. Together, these observations further confirmed the general recommendation for the prediction model to be trained on data comprising variations due to distinct parameters (such as the acquisition protocol and demographic characteristic). Despite the fact that this recommendation might sound trivial, it actually complements previous recommendations emphasizing the importance of sample size for good prediction performance (Varoquaux et al. 2012; Varoquaux et al. 2017), but further points out that, not the size per se matters, but the range of variations that are covered.

When applied in a clinical context, i.e. when evaluated on the ADNI dataset, the proposed framework not only showed good age-prediction for the healthy subjects but in particular also captured premature aging in the context of MCI and dementia as indicated by positive BrainAGE scores (Fig 8). More specifically, the dramatic atrophy of AD patients was reflected by a mean BrainAGE score of almost 10 years, which is comparable to the findings of a previous study conducted by Franke et al. (2010). The sensitivity of our framework to brain

structural changes in clinical populations was underscored by the likewise elevated BrainAGE for MCI participants, which was lower than for those with AD but still on average in the range of 5-6 years, i.e., above the MAE in the population-based samples (Davatzikos et al. 2008; Franke and Gaser 2012). In other words, our framework accurately ranked the HC, MCI and AD groups with regards to their clinical progression from healthy to demented (considering MCI as a transitional stage between normal aging and dementia; Petersen, 2010). However, statistically discriminating those individuals among MCI patients who will progress towards Alzheimer's disease is a challenging issue (Davatzikos et al. 2009; Petersen 2010; Gaser et al. 2013; Moradi et al. 2015). While a classification approach could be more powerful for such purpose than age prediction (Franke and Gaser 2014; Wang et al. 2016; Beheshti et al. 2017), the latter could be combined with the former to quantify deviations from normal aging trajectories across clinical stages.

4.8 Brain Age estimation using our framework

Overall, our results demonstrate that models trained on highly heterogeneous life span sample (MIXED) can predict the age of any unseen subject with a precision of 6 years (irrespective of approach i.e., either on a cross-validation on MIXED dataset or on an independent dataset). Given the broad age range of the training sample (18 to 81 years), a precision of 6 years can be considered as a good performance from the technical side. Importantly, all previous brain age prediction studies likewise reported a precision of approximately 5 – 6 years in the context of life-span samples. The relationship between GMV and the chronological age is modulated by many factors (both environmental and genetic factors (Burgmans et al. 2009; Giedd et al. 2010; Harada et al. 2013; Luders et al. 2016; Cole et al. 2017)). When aiming to identify the relationship between brain structural pattern and age, those factors may introduce noise obscuring the systematic effects of age on brain structure. In addition to these factors, inclusion of participant with certain characteristics (such as, participants in younger age with unidentified subclinical brain alterations, or older adults representing above-normal (i.e. “super healthy”) aged participants) might as well deviate the prediction model to capture the systematic effect of age on brain structure (Burgmans et al. 2009). Accordingly, these factors and the noise they introduce could account for the precision gap of 5 - 6 years in brain age prediction studies. That is, the limited precision of life-span age prediction may less relate to technical limitations but rather indicate that structural changes over a period of around 5 years are smaller than variations related to confounding factors that would represent “non-relevant” noise to the model. However, this hypothesis might not hold true for all life periods. For instance, one can observe dramatic age-related structural changes in childhood (cf. Erus et al. 2015) the late life periods (cf. the

higher precision of ~4 year MAE for 1000BRAINS or MIXED_55-90), while age-related grey matter changes could be relatively minor during early and middle adulthood (Schippling et al. 2017). Further examinations of these issues in future studies could provide better understanding of neurobiological aging. Nevertheless, in the scope of the present study, these confounding factors do not prevent our framework (combining OPNMF with sparse regression model) to accurately capture normal variations related to age and deviance from normal variations in clinical populations.

4.9 Identification of the regions influencing the prediction

Our results revealed that the regions involved in the age prediction model in 1000BRAINS were sparser than those underlying the prediction of age in MIXED. In contrast, most of the brain regions (representing 73% of the total grey matter volume) seemed to underpin the prediction when the model was trained in the MIXED dataset (which covers the adult lifespan with subjects between 18 to 81 years old). Put simply, the model cannot be consistently restricted to a few regions for inferring subjects' age when the cohort covers the adult lifespan. Such a pattern could argue for a more complex pattern of grey matter variations across the whole adult life span than in the later life periods. Previous studies have demonstrated that many different patterns of changes occur across the adult life span in grey matter volume with notably some regions showing monotonic decrease of GM and other showing a clear inverted U-shape grey matter volume (GMV)-age relationships or a "delayed decline" (Ziegler et al. 2012; Douaud et al. 2014; Ziegler et al. 2014). Furthermore, as aforementioned, several factors may induce brain structural variations in the young and middle-aged adult brain, such as life style and environmental factors (Miller et al. 2016) complicating the relationship between age and grey matter.

In our study, in addition to the regions highlighted for age prediction in older sample, some regions, such as the amygdala, and the superior parietal lobule further contributed to age prediction when the model was trained on the young and middle age adults (MIXED) dataset. Interestingly, the amygdala is one region where GMV has been found to increase with age in relatively younger samples (8 to 30 years old; (Ostby et al. 2009)) and some authors have noted no age-related GMV changes in the amygdala in older samples (Good et al. 2001). Furthermore, structural covariance of the amygdala (with other brain regions) is known to be modulated by several factors such as gender (Mechelli 2005). Thus, we could assume that, in a prediction model mixing genders, the amygdala could be selected as an indicator modulating the pattern of relationship between other brain regions and age, despite this region per se does not show a strong, linear and universal GMV decrease with

age. Accordingly, when examining the pattern of association between GMV and age, we observed a mild general linear decline of grey matter volume with age, but with a high variance across the subjects in the MIXED sample (see Fig S4) suggesting that very different age-related grey matter change patterns might be observed in this brain region. Such a pattern allows us to assume that the GMV in the amygdala, taken as isolated information, cannot significantly contribute to the age prediction, particularly in the case of older participants. In other words, we assume that the grey matter changes in the amygdala are diverse and occur mainly in the young and middle age adult lifespan rendering this specific region informative for predicting age in the whole adult life span sample in combination with information from other regions. However, on its own, this region would not be particularly informative for age prediction in older populations.

The superior parietal cortex is another example of regions contributing to the prediction analysis when the training sample consisted of young adults in addition to the older adults, but not when the model was trained on older adults only. Terribilli et al. (2011) conducted a study mainly focusing on young and middle age adults (18 to 50 years old), in which GMV of the lateral parietal cortex (i.e., supramarginal, angular and superior parietal cortex) exhibited a nonlinear relationship with age. The non-linear trend reported by the authors could be explained by a quadratic fit, that is, GMV followed a linear decline until the end of the fourth decade and then showed a mild increase. When examining the relationship between GMV and age in the superior parietal region in our study, we observed a similar trend (see Fig S4), in which the mean GMV of the superior parietal region showed a sharp decrease until 40 years of age, but less pronounced change with age in later life. Thus, despite the fact that prediction models in general, (specially LASSO regression models) are inherently linear, identification of GMV in the superior parietal cortex as relevant for age prediction converges with previous data demonstrating that structural changes in this region occur mainly in the first decades of adult life, but not in periods later in life. Thus, visually examining the pattern of associations between GMV and age in regions contributing to the prediction in MIXED suggest that some regions may be informative for their relatively systematic changes in the first period of adult life (such as the superior parietal cortex) while others regions could contribute by introducing complementary information (such as the amygdala) despite not exhibiting a clear linear relationship with age across the sample.

The pattern of regions consistently contributing to the prediction in the older sample appeared more spatially specific. Many of the regions highlighted by these analyses such as the hippocampus, the temporo-occipital

region and the medial superior frontal gyrus have been shown to be strongly affected by aging in the older life period and more specifically, to follow a strong linear decrease in this life period (after 40-50 years old; Raz et al. 2010a; Douaud et al. 2014). However, some other regions, such as the regions around the central sulcus are not known to show systematic change with age in later life period. Thus, the pattern of regions contributing to the predictions in 1000BRAINS suggests that when the training sample is restricted to older populations, the model can be restricted to a few regions, whose grey matter volumes is systematically affected by the aging process in the later life period, as well as other regions that might not appear particularly informative from a neurobiological point of view but complement the information conveyed by the former regions.

Interpreting the multivariate brain pattern weighting in the prediction is usually not recommended (e.g. (Haufe et al. 2014) since the prediction is underlined by the combination of several element/feature (i.e. voxels in a voxel-wise representation of the data and components in the present study) and that the individual elements on themselves, taken in isolation, may not convey any neurobiological relevant information. However, we would argue that the relationship between the brain and the predicted variable should not be kept as a conceptually locked black box, that is, the multivariate aspect of the prediction does not imply that we should not at least try to understand why the given pattern is relevant for the model. As a metaphor, if a model uses the variable “number of children” and “country” for predicting the age of a person, obviously the variable “country” on its own is not informative for predicting the age of a person, in contrast, the number of children is partly informative. Hence, examining the combination of those two variables for predicting the age of a person can provide us more insight by suggesting that the relationship between age and the number of children is modulated by cultural factors. Similarly, the pattern of relationship between grey matter volume and age is assumed to be modulated by several factors, but whose influence remained relatively poorly understood. However, the current framework promoting relatively localized component as relevant features could help to explore this issue in future studies (such as how the complex pattern of structural variations in the amygdala influenced by gender can contribute to age prediction in healthy adults).

5 Conclusions and practical considerations for future studies

In conclusion, our study, which evaluated OPNMF-based compression of VBM data for age prediction in two different healthy adult cohorts, opens several new perspectives. First, we demonstrated that OPNMF compression allows age prediction with a precision that is well comparable to that achieved following PCA

compression but yields substantially more interpretable results. It also outperformed an atlas-based approach based on resting-state whole-brain parcellation, even though the precision obtained by cross-model atlas based data compression is in itself remarkable. Considering the declining return of investment when going to higher granularities, we would thus suggest that OPNMF at a granularity of 300 and 500 components may provide the optimal data compression for age prediction.

While the exact OPNMF solution obviously depends on the examined sample, we here showed that prediction accuracies are basically uncompromised when employing a factorization derived from an independent dataset. That is, a factorization derived from one dataset can be used to efficiently compress VBM data of a second, independent dataset in a prediction framework. To note, the MIXED dataset used in the current study covers a wide range of variation over a broad age range while the (single-site) 1000BRAINS datasets can be assumed to capture structural covariance in older populations. Accordingly, the factorization derived from MIXED could be used for data compression in age prediction studies across adulthood whereas the factorization derived from 1000BRAINS may be particularly well suited for studying the aging brain. In addition to structural covariance-based factorization, our study offers robust prediction models trained on life span sample from heterogeneous sites (MIXED), an advantage on which future studies could capitalize to better understand the effects of different factors on the neurobiological aging.

ACKNOWLEDGEMENTS

This study was supported by the Deutsche Forschungsgemeinschaft (DFG, EI 816/4-1, GE 2835/1-1), the National Institute of Mental Health (R01-MH074457), the Helmholtz Portfolio Theme “Supercomputing and Modelling for the Human Brain”, the Helmholtz Initiative and Networking Fund (SC), and the European Union’s Horizon 2020 Research and Innovation Programme under Grant Agreement No. 7202070 (HBP SGA1), National Institute of Health (R01-AG014971) (AS & CD), National Institute on Aging (RF1-AG054409) (AS & CD).

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Ethical approval

The local ethics committees of the university hospital Aachen and Heinrich-Heine-University Düsseldorf have approved the original study protocol of the data used here. All the participants prior to the examination obtained informed consent. The current data were analyzed anonymously.

References

- Ashburner J (2007) A fast diffeomorphic image registration algorithm. *Neuroimage* 38:95–113. doi: 10.1016/j.neuroimage.2007.07.007
- Ashburner J, Yeung K, Ruzzo W, et al (1985) VBM Tutorial. *J Classif* 2:193–218. doi: 10.1007/BF01908075
- Beheshti I, Demirel H, Matsuda H (2017) Classification of Alzheimer's disease and prediction of mild cognitive impairment-to-Alzheimer's conversion from structural magnetic resonance imaging using feature ranking and a genetic algorithm. *Comput Biol Med* 83:109–119. doi: 10.1016/j.combiomed.2017.02.011
- Bellec P, Rosa-Neto P, Lyttelton OC, et al (2010) Multi-level bootstrap analysis of stable clusters in resting-state fMRI. *Neuroimage* 51:1126–1139. doi: 10.1016/j.neuroimage.2010.02.082
- Boutsidis C, Gallopoulos E (2008) SVD based initialization: A head start for nonnegative matrix factorization. *Pattern Recognit* 41:1350–1362. doi: 10.1016/j.patcog.2007.09.010
- Bunea F, She Y, Ombao H, et al (2011) Penalized least squares regression methods and applications to neuroimaging. *Neuroimage* 55:1519–1527. doi: 10.1016/j.neuroimage.2010.12.028
- Burgmans S, van Boxtel MPJ, Vuurman EFPM, et al (2009) The prevalence of cortical gray matter atrophy may be overestimated in the healthy aging brain. *Neuropsychology* 23:541–550. doi: 10.1037/a0016161
- Caspers S, Moebus S, Lux S, et al (2014) Studying variability in human brain aging in a population-based German cohort-rationale and design of 1000BRAINS. *Front Aging Neurosci* 6:1–14. doi: 10.3389/fnagi.2014.00149
- Cole JH, Poudel RPK, Tsagkrasoulis D, et al (2017) Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *Neuroimage* 163:115–124. doi: 10.1016/j.neuroimage.2017.07.059
- Craddock RC, James GA, Holtzheimer PE, et al (2012) A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Hum Brain Mapp* 33:1914–1928. doi: 10.1002/hbm.21333
- Davatzikos C (2016) Computational neuroanatomy using brain deformations: From brain parcellation to multivariate pattern analysis and machine learning. *Med. Image Anal.* 33:149–154.

- Davatzikos C (2004) Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. *Neuroimage* 23:17–20. doi: 10.1016/j.neuroimage.2004.05.010
- Davatzikos C, Fan Y, Wu X, et al (2008) Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging. *Neurobiol Aging* 29:514–523. doi: 10.1016/j.neurobiolaging.2006.11.010
- Davatzikos C, Xu F, An Y, et al (2009) Longitudinal progression of Alzheimers-like patterns of atrophy in normal older adults: The SPARE-AD index. *Brain* 132:2026–2035. doi: 10.1093/brain/awp091
- Douaud G, Groves AR, Tamnes CK, et al (2014) A common brain network links development, aging, and vulnerability to disease. *Proc Natl Acad Sci U S A* 111:17648–17653. doi: 10.1073/pnas.1410378111
- Eickhoff SB, Constable RT, Yeo BTT (2017) Topographic organization of the cerebral cortex and brain cartography. *Neuroimage*
- Erus G, Battapady H, Satterthwaite TD, et al (2015) Imaging patterns of brain development and their relationship to cognition. *Cereb Cortex* 25:1676–1684. doi: 10.1093/cercor/bht425
- Finn ES, Shen X, Scheinost D, et al (2015) Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat Neurosci* 18:1–11. doi: 10.1038/nn.4135
- Franke K, Gaser C (2012) Longitudinal changes in individual BrainAGE in healthy aging, mild cognitive impairment, and Alzheimer's disease. *GeroPsych J Gerontopsychology Geriatr Psychiatry* 25:235–245. doi: 10.1024/1662-9647/a000074
- Franke K, Gaser C (2014) Dementia classification based on brain age estimation. *Proc MICCAI Work Chall Comput Diagnosis Dement Based Struct MRI Data* 48–54.
- Franke K, Gaser C, Manor B, Novak V (2013) Advanced BrainAGE in older adults with type 2 diabetes mellitus. *Front Aging Neurosci* 5:1–9. doi: 10.3389/fnagi.2013.00090
- Franke K, Luders E, May A, et al (2012) Brain maturation: Predicting individual BrainAGE in children and adolescents using structural MRI. *Neuroimage* 63:1305–1312. doi: 10.1016/j.neuroimage.2012.08.001
- Franke K, Ristow M, Gaser C (2014) Gender-specific impact of personal health parameters on individual brain aging in cognitively unimpaired elderly subjects. *Front Aging Neurosci* 6:1–14. doi: 10.3389/fnagi.2014.00094
- Franke K, Ziegler G, Klöppel S, Gaser C (2010) Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: Exploring the influence of various parameters. *Neuroimage* 50:883–892. doi: 10.1016/j.neuroimage.2010.01.005
- Gaser C, Franke K, Klöppel S, et al (2013) BrainAGE in Mild Cognitive Impaired Patients: Predicting the

- Conversion to Alzheimer's Disease. PLoS One. doi: 10.1371/journal.pone.0067346
- Genon S, Li H, Fan L, et al (2017) The Right Dorsal Premotor Mosaic: Organization, Functions, and Connectivity. *Cereb Cortex* 27:2095–2110. doi: 10.1093/cercor/bhw065
- Genon S, Reid A, Li H, et al (2016) The heterogeneity of the left dorsal premotor cortex evidenced by multimodal connectivity-based parcellation and functional characterization. *Neuroimage*
- Giedd JN, Stockman M, Weddle C, et al (2010) Anatomic magnetic resonance imaging of the developing child and adolescent brain and effects of genetic variation. *Neuropsychol. Rev.* 20:349–361.
- Glasser MF, Coalson TS, Robinson EC, et al (2016) A multi-modal parcellation of human cerebral cortex. *Nature* 536:171–178. doi: 10.1038/nature18933
- Good CD, Johnsrude IS, Ashburner J, et al (2001) A voxel-based morphometric study of ageing in 465 normal adult human brains. *Neuroimage* 14:21–36. doi: 10.1006/nimg.2001.0786
- Gordon EM, Laumann TO, Adeyemo B, et al (2016) Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. *Cereb Cortex* 26:288–303. doi: 10.1093/cercor/bhu239
- Guyon I, Elisseeff A (2003) An Introduction to Variable and Feature Selection. *J Mach Learn Res* 3:1157–1182. doi: 10.1016/j.aca.2011.07.027
- Harada CN, Natelson Love MC, Triebel KL (2013) Normal cognitive aging. *Clin. Geriatr. Med.* 29:737–752.
- Haufe S, Meinecke F, G??rgen K, et al (2014) On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage* 87:96–110. doi: 10.1016/j.neuroimage.2013.10.067
- Hua J, Tembe WD, Dougherty ER (2009) Performance of feature-selection methods in the classification of high-dimension data. *Pattern Recognit* 42:409–424. doi: 10.1016/j.patcog.2008.08.001
- Hubert L, Arabie P (1985) Comparing partitions. *J Classif* 2:193–218. doi: 10.1007/BF01908075
- Jolliffe IT (2002) *Principal Component Analysis*, Second Edition.
- Kampa K, Mehta S, Chou CA, et al (2014) Sparse optimization in feature selection: Application in neuroimaging. *J Glob Optim* 59:439–457. doi: 10.1007/s10898-013-0134-2
- Kelly C, Toro R, Di Martino A, et al (2012) A convergent functional architecture of the insula emerges across imaging modalities. *Neuroimage* 61:1129–1142. doi: 10.1016/j.neuroimage.2012.03.021
- Koutsouleris N, Davatzikos C, Borgwardt S, et al (2014) Accelerated brain aging in schizophrenia and beyond: A neuroanatomical marker of psychiatric disorders. *Schizophr Bull* 40:1140–1153. doi: 10.1093/schbul/sbt142
- Lakkaraju H, Bach SH, Jure L (2016) Interpretable Decision Sets: A Joint Framework for Description and

- Prediction. *Kdd* 2016:1675–1684. doi: 10.1145/2939672.2939874
- Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. *Nature* 401:788–91. doi: 10.1038/44565
- Liem F, Varoquaux G, Kynast J, et al (2017) Predicting brain-age from multimodal imaging data captures cognitive impairment. *Neuroimage* 148:179–188. doi: 10.1016/j.neuroimage.2016.11.005
- Liu X, Yan S, Jin H (2010) Projective nonnegative graph embedding. *IEEE Trans Image Process* 19:1126–1137. doi: 10.1109/TIP.2009.2039050
- Luders E, Cherbuin N, Gaser C (2016) Estimating brain age using high-resolution pattern recognition: Younger brains in long-term meditation practitioners. *Neuroimage* 134:508–513. doi: 10.1016/j.neuroimage.2016.04.007
- May A (2011) Experience-dependent structural plasticity in the adult human brain. *Trends Cogn. Sci.* 15:475–482.
- Mechelli A (2005) Structural Covariance in the Human Cortex. *J Neurosci* 25:8303–8310. doi: 10.1523/JNEUROSCI.0357-05.2005
- Miller KL, Alfaro-Almagro F, Bangerter NK, et al (2016) Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat Neurosci* 19:1523–1536. doi: 10.1038/nn.4393
- Moradi E, Pepe A, Gaser C, et al (2015) Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *Neuroimage* 104:398–412. doi: 10.1016/j.neuroimage.2014.10.002
- Mwangi B, Hasan KM, Soares JC (2013) Prediction of individual subject's age across the human lifespan using diffusion tensor imaging: A machine learning approach. *Neuroimage* 75:58–67. doi: 10.1016/j.neuroimage.2013.02.055
- Mwangi B, Tian TS, Soares JC (2014) A review of feature reduction techniques in Neuroimaging. *Neuroinformatics* 12:229–244.
- Ostby Y, Tamnes CK, Fjell AM, et al (2009) Heterogeneity in subcortical brain development: A structural magnetic resonance imaging study of brain maturation from 8 to 30 years. *J Neurosci* 29:11772–11782. doi: 10.1523/JNEUROSCI.1242-09.2009
- Petersen RC (2010) Alzheimer's disease: progress in prediction. *Lancet Neurol* 9:4–5. doi: 10.1016/S1474-4422(09)70330-8
- Raz N, Ghisletta P, Rodrigue KM, et al (2010) Trajectories of brain aging in middle-aged and older adults: Regional and individual differences. *Neuroimage* 51:501–511. doi: 10.1016/j.neuroimage.2010.03.020

- Santos JM, Embrechts M (2009) On the use of the adjusted rand index as a metric for evaluating supervised classification. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). pp 175–184
- Schaefer A, Kong R, Gordon EM, et al (2017) Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cereb Cortex* 1–20. doi: 10.1093/cercor/bhx179
- Schippling S, Ostwaldt AC, Suppa P, et al (2017) Global and regional annual brain volume loss rates in physiological aging. *J Neurol* 264:520–528. doi: 10.1007/s00415-016-8374-y
- Sotiras A, Resnick SM, Davatzikos C (2015) Finding imaging patterns of structural covariance via Non-Negative Matrix Factorization. *Neuroimage* 108:1–16. doi: 10.1016/j.neuroimage.2014.11.045
- Sotiras A, Toledo JB, Gur RE, et al (2017) Patterns of coordinated cortical remodeling during adolescence and their associations with functional specialization and evolutionary expansion. *Proc Natl Acad Sci* 114:3527–3532. doi: 10.1073/pnas.1620928114
- Sporns O (2013) Network attributes for segregation and integration in the human brain. *Curr. Opin. Neurobiol.* 23:162–171.
- Terribilli D, Schaefelberger MS, Duran FLS, et al (2011) Age-related gray matter volume changes in the brain during non-elderly adulthood. *Neurobiol Aging* 32:354–368. doi: 10.1016/j.neurobiolaging.2009.02.008
- Thirion B, Varoquaux G, Dohmatob E, Poline JB (2014) Which fMRI clustering gives good brain parcellations? *Front Neurosci* 8:1–13. doi: 10.3389/fnins.2014.00167
- Tibshirani R (1996) Regression Selection and Shrinkage via the Lasso. *J. R. Stat. Soc. B* 58:267–288.
- Tipping M (2001) Sparse Bayesian Learning and the Relevance Vector Mach. *J Mach Learn Res* 1:211–244. doi: 10.1162/15324430152748236
- Tipping ME, Faul AC (2003) Fast Marginal Likelihood Maximisation for Sparse Bayesian Models. Ninth Int Work Artificial Intell Stat 1–13. doi: 10.1.1.165.4281
- Tisserand DJ, Pruessner JC, Sanz Arigita EJ, et al (2002) Regional frontal cortical volumes decrease differentially in aging: An MRI study to compare volumetric approaches and voxel-based morphometry. *Neuroimage* 17:657–669. doi: 10.1016/S1053-8119(02)91173-0
- Tucholka A, Thirion B, Perrot M, et al (2008) Probabilistic anatomo-functional parcellation of the cortex: How many regions? In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). pp 399–406
- Varoquaux G, Gramfort A, Thirion B (2012) Small-sample brain mapping: sparse recovery on spatially

- correlated designs with randomization and clustering. *Proc 29th Int Conf Mach Learn* 1375–1382.
- Varoquaux G, Raamana PR, Engemann DA, et al (2017) Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. *Neuroimage* 145:166–179. doi: 10.1016/j.neuroimage.2016.10.038
- Wang P, Chen K, Yao L, et al (2016) Multimodal classification of mild cognitive impairment based on partial least squares. *J Alzheimer's Dis* 54:359–371. doi: 10.3233/JAD-160102
- Yang Z, Oja E (2010) Linear and Nonlinear Projective Nonnegative Matrix Factorization. *Trans Neur Netw* 21:734–749. doi: 10.1109/TNN.2010.2041361
- Yang Z, Yuan Z, Laaksonen J (2007) Projective Non-negative matrix factorization with application to facial image processing. *Pattern Recognit* 21:1353–1362.
- Yuan Z, Yang Z, Oja E (2007) Projective Nonnegative Matrix Factorization : Sparseness , Orthogonality , and Clustering. *Technology* 1–14.
- Zhang CH, Huang J (2008) The sparsity and bias of the lasso selection in high-dimensional linear regression. *Ann Stat* 36:1567–1594. doi: 10.1214/07-AOS520
- Zhao P, Yu B (2006) On Model Selection Consistency of Lasso. *J Mach Learn Res* 7:2541–2563. doi: 10.1109/TIT.2006.883611
- Ziegler G, Dahnke R, Jäncke L, et al (2012) Brain structural trajectories over the adult lifespan. *Hum Brain Mapp* 33:2377–2389. doi: 10.1002/hbm.21374
- Ziegler G, Ridgway GR, Dahnke R, Gaser C (2014) Individualized Gaussian process-based prediction and detection of local and global gray matter abnormalities in elderly subjects. *Neuroimage* 97:333–348. doi: 10.1016/j.neuroimage.2014.04.018
- Zou H, Hastie T (2005) Regularization and variable selection via the elastic-net. *J R Stat Soc* 67:301–320. doi: 10.1111/j.1467-9868.2005.00503.x

Figure Legends

Fig 1. The two healthy datasets and the non-negative matrix components derived the these datasets. A: Overview of the sample characteristics of the two datasets (i.e., range of age distributed in each dataset, as well as the scanner protocol). B: Brain spatial representation of the factorization derived from the two datasets at two different resolutions. C: Similarity between the factorizations derived from the two datasets.

Fig 2. Different prediction approaches evaluated in the study. A&B illustrate the procedure of a 10 fold cross-validation when compressing a dataset using the components derived from itself (A) MIXED dataset (B) 1000BRAINS. C&D illustrate the procedure of a 10 fold cross-validation performed on features extracted by using the components derived from the other dataset (C) MIXED dataset projected on to 1000BRAINS based factorization, and (D) 1000BRAINS projected on to factorization derived from MIXED. E,F&G illustrate the approaches utilizing an independent dataset to validate the model trained on the dataset compressed using the components derived from itself (E) training the model on 1000BRAINS dataset projected on to its own factorization and later validate the model on OldMIXED dataset, (F) training the model on OldMIXED dataset projected on to OldMIXED based factorization and later validate the model on 1000BRAINS and (G) training the model on MIXED dataset projected on to MIXED based factorization and later validate the model on 1000BRAINS.

Fig 3. Main processing steps for age prediction based on GMV and the post-hoc examination of regions contributing to the prediction. 1) Voxel-based morphometric (VBM) data for each subject are used as input for OPNMF 2) Following OPNMF, the VBM data are represented by two matrices, denoting the corresponding subject-specific loading coefficients (H) and the sparse components (W). 3) Application of sparse regression model, in which H provides the features for the prediction model 4) Evaluation of the prediction model using a test sample (different prediction models described in section 2.5 & Fig 2). 5&6) Identification of the regions contributing in the prediction analysis; 5) First the respective components with non-zero coefficients assigned by the prediction models were identified. Then, we built a “contributor map” at each level of granularity, in which the voxel values represent their (binary) contribution in at least 95% of the models. 6) Combining those maps (by summing up the values) resulted in a contributor “summary” map in which a non-zero value represents a contribution in at least 95% of all the prediction models) and higher value represent higher overlap across different granularities. As our analyses revealed that prediction performance stabilizes around 300 components (Fig 6), only the contributor maps of granularity > 300 components were merged into the summary map.

Fig 4. Chronological age plotted against the age predicted using the high-dimensional VBM data. The lower figure exhibits the isolated voxels that contributed in the prediction analysis. Here, the voxels, which contributed in all the models across 25 replications of 3-fold cross-validation, are displayed.

Fig 5. Illustration of the mean absolute error of the prediction models using different compression methods with various sparse regression models (LASSO+OPNMF in black; LASSO+PCA in blue; RVM+OPNMF in red; RVM+PCA in magenta) at different levels of granularity, and separately for each prediction approach. A) represents the approach where 10 fold cross-validation is performed and B) illustrates the approaches in which an independent dataset is used to validate the model trained on the dataset compressed based on the components derived from itself.

Fig 6. Illustration of the mean absolute error of the prediction models using different spatially localized compression models (OPNMF and RS-parcellation) with sparse (LASSO) regression model at different levels of granularity, and separately for each prediction approach. A) represents the approach where 10 fold cross-validation is performed and B) illustrates the approaches in which an independent dataset is used to validate the model trained on the dataset compressed based on components derived from itself.

Fig 7. Chronological age plotted against the age predicted using the VBM data compressed with OPNMF. The predicted age plotted in this figure is an average of the predicted age across different levels of granularity.

Fig 8. Validation approach in ADNI samples. BrainAGE scores (reflecting the difference between predicted age and the chronological age) are showed for all the three subsamples (i.e., Healthy controls (HC), Mild Cognitive Impairment (MCI), Alzheimer's Disease (AD)). The left plot refers to the approach in which the model was trained on OldMIXED sample compressed using factorization derived from the whole MIXED sample. The right plot refers to the approach in which the model was trained on 1000BRAINS compressed using factorization derived from 1000BRAINS.

Fig 9. Summary map of the regions that contributed in the prediction analysis when performing 10-fold cross-validation and compressing the dataset using the components derived from the other dataset, in which brighter shade represents more frequently used parts of the brain. Plain anatomical slices are displayed as reference in the top row. The middle row illustrates the MIXED dataset compressed with the 1000BRAINS-based factorization while the bottom row illustrates 1000BRAINS dataset compressed with the MIXED-based factorization.

